Does Constructing A Belief Distribution Truly Reduce Overconfidence?

Beidi Hu and Joseph P. Simmons

The Wharton School, University of Pennsylvania

Forthcoming in the *Journal of Experimental Psychology: General*

Author Note

We thank the Wharton Behavioral Lab, the Wharton Dean's Fund, and the Wharton Risk Management Center's Russell Ackoff Doctoral Student Fellowship for financial support; attendees at Deborah Small's lab meeting, seminar participants at NYU Stern, Don Moore, and the review team for valuable feedback; and Jake Flancer and Oliver Zou for research assistance. Some of the present research was presented at the conference on Subjective Probability, Utility, and Decision Making (August 2021), at the annual convention of the Society for Personality and Social Psychology (February 2022), and at the annual meeting of the Society for Judgment and Decision Making (February 2022).

Correspondence concerning this article should be addressed to Beidi Hu, Department of Operations, Information, and Decisions, The Wharton School, University of Pennsylvania, 500 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104, United States. Email: beidihu@wharton.upenn.edu

All of our data, materials, and pre-registrations are available at: https://researchbox.org/314.

Abstract

Can overconfidence be reduced by asking people to provide a belief distribution over all possible outcomes – that is, by asking them to indicate how likely *all* possible outcomes are? Although prior research suggests that the answer is "yes," that research suffers from methodological confounds that muddle its interpretation. In our research, we remove these confounds to investigate whether providing a belief distribution truly reduces overconfidence. In 10 studies, participants made predictions about upcoming sports games or other participants' preferences, and then indicated their confidence in these predictions using rating scales, likelihood judgments, and/or incentivized wagers. Contrary to prior research, and to our own expectations, we find that providing a belief distribution usually *increases* overconfidence, because doing so seems to reinforce people's prior beliefs.

Keywords: overconfidence, belief distribution, judgment under uncertainty, debiasing

People often have too much confidence in the accuracy of their forecasts and beliefs (e.g., Alpert & Raiffa, 1982; Block & Harper, 1991; Klayman et al., 1999; Lichtenstein et al., 1977; Soll & Klayman, 2004; Teigen & Jørgensen, 2005). This form of overconfidence, dubbed *overprecision* (Moore & Healy, 2008), can have dire consequences, as when people fail to anticipate, plan for, or prevent disasters that were erroneously deemed impossible (e.g., Higginbotham, 2019; Lewis, 2011; McLean & Elkind, 2013).

Not surprisingly, researchers have long been interested in identifying ways to reduce overprecision (e.g., Arkes et al., 1987; Juslin et al., 1999; Koriat et al., 1980; Lichtenstein & Fischhoff, 1980; Soll & Klayman, 2004; Teigen & Jørgensen, 2005; Walters et al., 2017). One recent and promising suggestion involves asking people to provide belief distributions over all possible outcomes, rather than merely asking them for point estimates or interval estimates (Haran et al., 2010; Moore, 2020). Figure 1 presents an example of how such belief distributions can be elicited, using an interface developed by Haran et al. (2010). Essentially, it involves presenting people with the entire range of possible outcomes (e.g., 0% to 100%), partitioned into mutually exclusive and collectively exhaustive intervals, and asking them to indicate the probability that each interval includes the correct answer. Researchers in psychology and related fields are increasingly eliciting belief distributions as an attempt to acquire a more complete understanding of participants' beliefs (André et al., 2022; Dietvorst & Bharti, 2020; Goldstein & Rothschild, 2014; Hofman et al., 2020; Moore et al., 2015; Moore et al., 2017; Prims & Moore, 2017; Reinholtz et al., 2021; Ren & Croson, 2013; Soll et al., 2019).

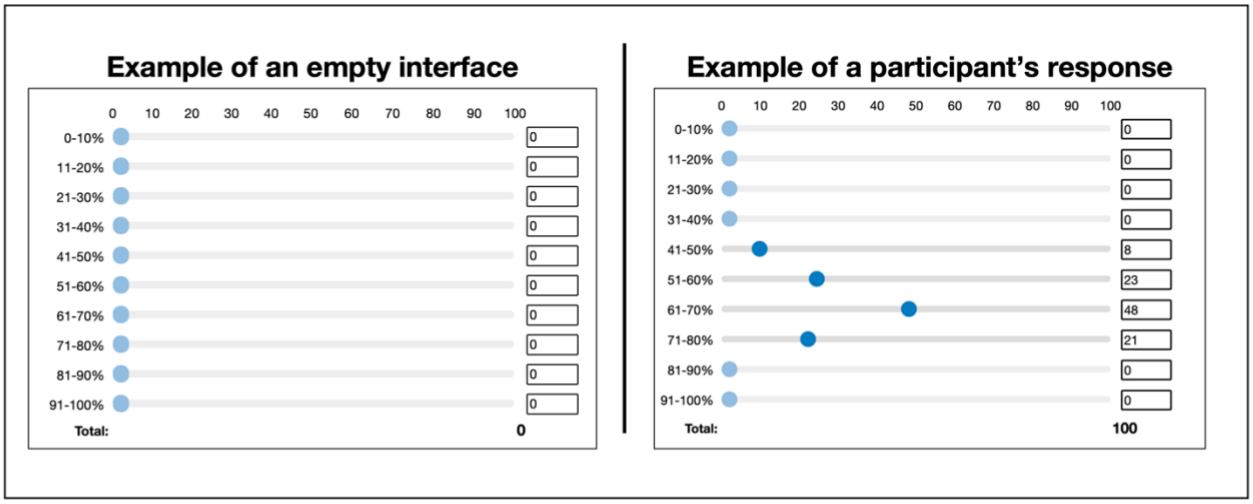

Figure 1. An Example of the Belief Distribution Interface

Note. The interface contains an array of sliders, each representing an outcome interval within the entire range of possible outcomes (e.g., 0% to 100%). Participants use sliders to indicate the probability that the correct answer falls within each interval. The right panel presents an example response of a participant who indicates that there is an 8% chance that the outcome is between 41-50%, a 23% chance that it is between 51-60%, a 48% chance that it is between 61-70%, and a 21% chance that it is between 71-80%.

Prior research suggests that, compared with the traditional way of eliciting confidence intervals, eliciting belief distributions "effectively reduces overprecision" and sometimes "completely eliminates overprecision," as evidenced by participants producing wider confidence intervals, which are in turn more likely to contain the correct answer (Haran et al., 2010). Why would asking people to provide a belief distribution reduce overprecision? As Moore (2020) writes in his excellent book on overconfidence, "Asking people to complete a [belief distribution] . . . forces them to broaden their thinking and consider the possibility that their best guess is wrong" (p. 69; see also Moore, 2022). In other words, by forcing people to consider a wider range of possible outcomes – including those that would not have been considered otherwise – it encourages them to realize that some outcomes are more likely than they would have otherwise thought, thereby increasing the width of their confidence intervals and reducing overconfidence.

This rationale for why providing a belief distribution would reduce overconfidence makes good sense and is consistent with the literature on the "considering the alternative" approach. People tend to overweigh evidence in favor of their beliefs relative to alternative outcomes (Hoch, 1985; Klayman, 1995; Klayman & Ha, 1987; Koriat et al., 1980). Most successful attempts to mitigate this tendency have taken the form of "considering the alternative," which prompts decision makers to consider supporting evidence for other possible outcomes. One way to implement this strategy is through direct instructions. Past research that explicitly asks participants to consider alternative or unknown possibilities has effectively reduced the tendency to interpret evidence in favor of prior beliefs (Lord et al., 1984), increased judgment accuracy (Hoch, 1985; Williams & Mandel, 2007), and reduced overconfidence (Koriat et al., 1980; Walters et al., 2017). A second approach is to alter the elicitation format so as to indirectly prompt people to consider the alternative. For example, Soll and Klayman (2004) found that separately asking for the lower and upper bounds of a confidence interval slightly mitigated overconfidence compared with asking for a single range estimate. They suggest that separately considering the two points requires sampling one's knowledge twice and encourages retrieval of a wider range of evidence (see also Speirs-Bridge et al., 2010; Teigen & Jørgensen, 2005).

Similar reasoning underlies the belief distribution method as a remedy for overprecision. Eliciting a belief distribution forces people to consider the entire range of possible outcomes, including unlikely outcomes that would not normally come to mind. It is extremely reasonable to expect that, like the successful "consider the alternative" interventions described above, the belief distribution elicitation would reduce people's overconfidence. Furthermore, whereas prior attempts of "considering the alternative" usually involve heavy-handed instructions that might lead to experimenter demand, providing a belief distribution works in a subtler way and might be more successful for real-world applications.

Nevertheless, the best evidence supporting the effects of providing belief distributions on overconfidence suffers from a potentially important methodological shortcoming. In past research, participants have been randomly assigned to either provide a belief distribution or to directly provide a 90% confidence interval (Haran et al., 2010). Researchers have found that 90% confidence intervals implied by participants' belief distributions are wider (and more likely to include the correct answer) than 90% confidence intervals that are directly provided by participants. Although this suggests that belief distributions may reduce overprecision and increase calibration, there is an alternative and potentially artifactual explanation that needs to be ruled out. Specifically, it could be that when participants encounter belief distributions, they feel compelled by the nature of the task to assign some probability to most or all of the provided intervals, even if they do not actually believe that the interval's true probability is greater than zero. In other words, some participants may feel like they *should* assign some probability to some outcomes even if they think that these outcomes are impossible (see Figure 2).[1] Even if only a subset of participants were susceptible to this form of task demand, it would artificially widen the intervals imputed from those participants' belief distributions. Thus, what appears to be a remedy for overconfidence may instead be a methodological artifact, and one cannot definitively conclude that those providing belief distributions were indeed less overconfident.

[1] Our studies were not designed to directly test this possibility. Nevertheless, we did find evidence consistent with the task demand explanation in Studies 9 and 10 (see Footnote 10).

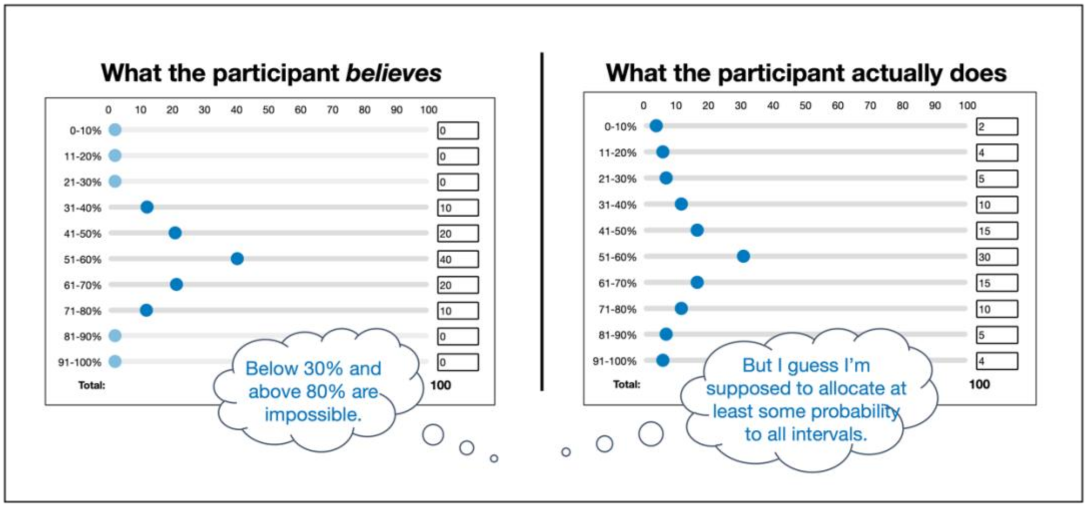

Figure 2. Example of How Belief Distribution Elicitations May Not Elicit True Beliefs

In our research, we sought to provide a clean test of the effect of providing belief distributions on overprecision. To achieve this, we (1) asked participants to provide a best guess about some outcome, (2) randomly assigned them to provide a belief distribution or not, and then (3) assessed participants' confidence in their best guess, using measures of self-reported confidence, likelihood estimates, and/or incentive-compatible wager tasks. By measuring confidence the same way in all conditions, we effectively eliminate the measurement confounds that potentially plagued previous research.[2] If providing belief distributions truly reduces (over)confidence, then participants who do so should be less confident in their best guesses, report them as being less likely to be correct, and be less willing to bet on their accuracy. Of note, we did not measure overconfidence by eliciting confidence intervals, not only because of its potential in confounding the results in this context (as noted in the preceding paragraph), but also because, in general, we do not think it is

[2] Haran et al. (2010) sidestep these confounds in their Study 3, by finding that participants' 90% confidence intervals were wider for some items after having provided belief distributions for previous items. We conducted a cleaner test of this hypothesis in our Study S1 (described in detail in Supplement 8) and were unable to replicate it.

ideal to assess overconfidence with confidence intervals, as these intervals tend to have very poor measurement properties (e.g., Langnickel & Zeisberger, 2016; Teigen & Jørgensen, 2005).[3]

We went into this research expecting these cleaner tests to confirm what previous research had found: that providing belief distributions reduces overconfidence. Surprisingly (to us, at least), we tended to find the opposite. In 10 pre-registered studies, we found that providing belief distributions sometimes had no effect but usually *increased* confidence. It seems that providing a belief distribution often convinces people that they were, in fact, right all along.

Research Overview

In this paper, we present 10 pre-registered experiments investigating whether providing a belief distribution truly reduces (over)confidence. To prevent online participants from looking up the answers, we asked participants to predict inherently unknown quantities. In some studies, participants predicted the outcomes of upcoming sports games; in others, they predicted the preferences of other survey respondents. To remove the methodological confounds present in past research, we measured confidence the same way across all conditions. Table 1 presents the measures we collected in each study.

[3] Indeed, in our Study 2, we find that the measures we use to assess overconfidence in our studies are much more highly correlated with a behavioral measure of overconfidence (i.e., willingness to wager on one's beliefs; *r*s = .28, *p*s < .001) than either directly elicited 90% confidence interval width ($r = -.05$, $p = .226$) or 90% confidence interval width imputed from participants' belief distributions ($r = -.11$, $p = .013$).

Table 1. Studies 1-10: Confidence Measures

Study 1
How confident are you that [your predicted winner] will win? *(1 = Not at all confident, 9 = Extremely confident)*
In your opinion, how likely are [your predicted winner] to win? *(0% to 100%)*
Studies 2, 9, 10
How confident are you that your prediction is within 5 points of what they will score in this game? *(1 = Not at all confident, 9 = Extremely confident)*
In your opinion, how likely are [the team] to score within 5 points of what you predicted? *(0% to 100%)*
Study 2 only: What do you want to do? *(0 = Not wager, and receive an additional 20 cents; 1 = Wager, for the opportunity to receive an additional 40 cents)*
Studies 3, 4 (low precision condition), 5, 7, 8
How confident are you that your prediction is within 5 percentage points of the correct answer? *(1 = Not at all confident, 9 = Extremely confident)*
In your opinion, how likely is the correct answer within 5 percentage points of what you predicted? *(0% to 100%)*
Study 4 (high precision condition)
How confident are you that your prediction is correct? *(1 = Not at all confident, 9 = Extremely confident)*
In your opinion, how likely is your prediction correct? *(0% to 100%)*
Study 6
How confident are you that your prediction is within 3 points of what they will score in this game? *(1 = Not at all confident, 9 = Extremely confident)*
In your opinion, how likely are [the team] to score within 3 points of what you predicted? *(0% to 100%)*

In Studies 1 and 2, we provide basic tests of the effects of providing a belief distribution on overconfidence. Contrary to what prior research suggests, we found that providing a belief distribution does not reduce overconfidence, and sometimes increases it. In Studies 3-5, we examine the robustness and generalizability of this result, finding that it does not depend on whether belief distributions are elicited before or after participants provide their best guesses (Study 3), the precision of the prediction question (Study 4), or whether the question's correct answer is extreme or moderate (Study 5). In Studies 6-8, we test several interventions aimed at reducing people's confidence, all of which either backfired or were ineffectual. Nevertheless, the results of these studies suggested a possible mechanism that we examined in Studies 9 and 10. In these studies, we found that providing belief distributions increases (over)confidence because the act of allocating probabilities to outcomes seems to reinforce people's existing beliefs.

We report all of our measures, manipulations, and exclusions; all of our sample sizes were determined in advance. A detailed breakdown of all exclusions for all studies can be found in Supplement 1. All of our data, materials, and pre-registrations are available on ResearchBox: https://researchbox.org/314. This research was approved by the University of Pennsylvania's Institutional Review Board (Protocol 834437).

Studies 1 and 2

In Studies 1 and 2, we aimed to provide clean tests of the effects of providing a belief distribution on confidence. We asked participants to predict the outcomes of upcoming National Football League (NFL) games, and we assessed participants' confidence the same way in all conditions: using confidence rating scales, likelihood estimates, and, in Study 2, an incentivized wager task. After making predictions and before completing the confidence measures, participants were randomly assigned to provide a belief distribution, a 90% confidence interval, or, in Study 2, neither.

Method

Participants. We conducted Studies 1 and 2 using U.S. participants from Prolific. We decided in advance to recruit 600 participants and 1,000 participants, respectively.

In Study 1, we pre-registered to retain only the first response from Prolific IDs or IP addresses that appeared more than once in our dataset (21 exclusions) and to exclude participants who misreported Prolific IDs (6 exclusions) and participants who failed the attention check (66 exclusions).[4] This left us with a final sample of 511, which averaged 25.3 years of age and was 35.6% female.

In Study 2, we pre-registered to exclude all responses from duplicate Prolific IDs or duplicate IP addresses (44 exclusions), from participants who misreported Prolific IDs (4 exclusions), and from participants who failed the attention check (132 exclusions). We also pre-registered to only

[4] In studies with sports predictions (Studies 1, 2, 6, 9, and 10), we asked participants an attention check question at the end of the survey. In Study 2 (in which participants made predictions for two games), we asked participants to choose the two games they predicted. In Studies 1, 6, 9, and 10 (in which participants made predictions for four games), we asked participants to choose the game that they did NOT predict. Per our pre-registration, we manually excluded participants who failed this attention check question. In studies with preference/behavior predictions (Studies 3-5, 7-8, S1, and S3), participants first answered the set of preference/behavior questions for themselves. We then asked them to choose the question they did NOT respond to. Participants who failed this attention check question were automatically excluded from the survey.

allow participants who self-reported in the pre-screening questions to have watched an entire NFL football game and identified themselves as NFL fans to proceed with the survey. However, due to restrictions in Prolific policy, we were forced to allow non-NFL fans to proceed with the survey. The final sample for Study 2 was 812, which averaged 35.8 years of age and was 37.9% female. Among those, 583 participants met the pre-screening criteria (i.e., self-reported to have watched an entire NFL game and identified themselves as NFL fans). To be consistent with our pre-registration, we discuss the results below only including those 583 fans; in Supplement 2 we report the results including all 812 participants. A detailed breakdown of all exclusions for all studies can be found in Supplement 1.

Procedures. In each study, participants were asked to predict the outcomes of several upcoming NFL games. In Study 1, which was conducted on October 9, 2020, participants predicted which team would win four games played on October 11, 2020. In Study 2, which was conducted on November 20-21, 2020, participants predicted the total points scored by a team for two games, randomly selected from a set of four games played on November 22, 2020. (See Supplement 3 for the full list of games.) For each game, participants saw the starting time, the name of the home team and the visiting team, and the current win-loss records for each team. The games were presented in a random order.

For each game, participants first made a prediction. In Study 1, they indicated which team would win. In Study 2, they provided their best estimate of how many points would be scored by a randomly selected team (i.e., either the home team or the visiting team). After making their predictions, participants moved on to the next screen of the survey, at which point they were randomly assigned to provide a belief distribution (Belief Distribution condition), a 90% confidence interval (Confidence Interval condition), or neither of these (Control condition, only in

Study 2). In Study 1, participants provided the belief distribution or confidence interval for the point differential of the game. In Study 2, participants who were not in the Control condition provided the belief distribution or confidence interval for the points scored by the selected team.

In the Belief Distribution condition, we elicited belief distributions by asking participants to assign probabilities to each of eight outcome intervals. In Study 1, the labels for the intervals read: [The team they picked as the winning team] will "lose by more than 30 points," "lose by 21 to 30 points," "lose by 11 to 20 points," "lose by 1 to 10 points," "tie or win by 1 to 9 points," "win by 10 to 19 points," "win by 20 to 29 points," and "win by 30 points or more." In Study 2, the labels for the intervals read: [The team] will score "6 points or fewer," "7 to 12 points," "13 to 18 points," "19 to 24 points," "25 to 30 points," "31 to 36 points," "37 to 42 points," and "43 points or more." The survey software forced the probabilities to add up to 100 across all eight categories. Figure 3 shows what the confidence interval and belief distribution elicitations looked like in Study 2.

Then, on a separate page, we measured participants' confidence in their predictions using a confidence rating question and a likelihood estimate question (see Table 1). In Study 2, we also included an incentive-compatible wager measure. After responding to the confidence rating and likelihood estimate questions, participants received a bonus of 20 cents. They were asked whether they would like to wager the additional bonus on their predictions being within 5 points of the true outcome. If they said "yes", then their bonus would double if the team's actual score was within 5 points of their prediction; otherwise, they would lose their bonus.

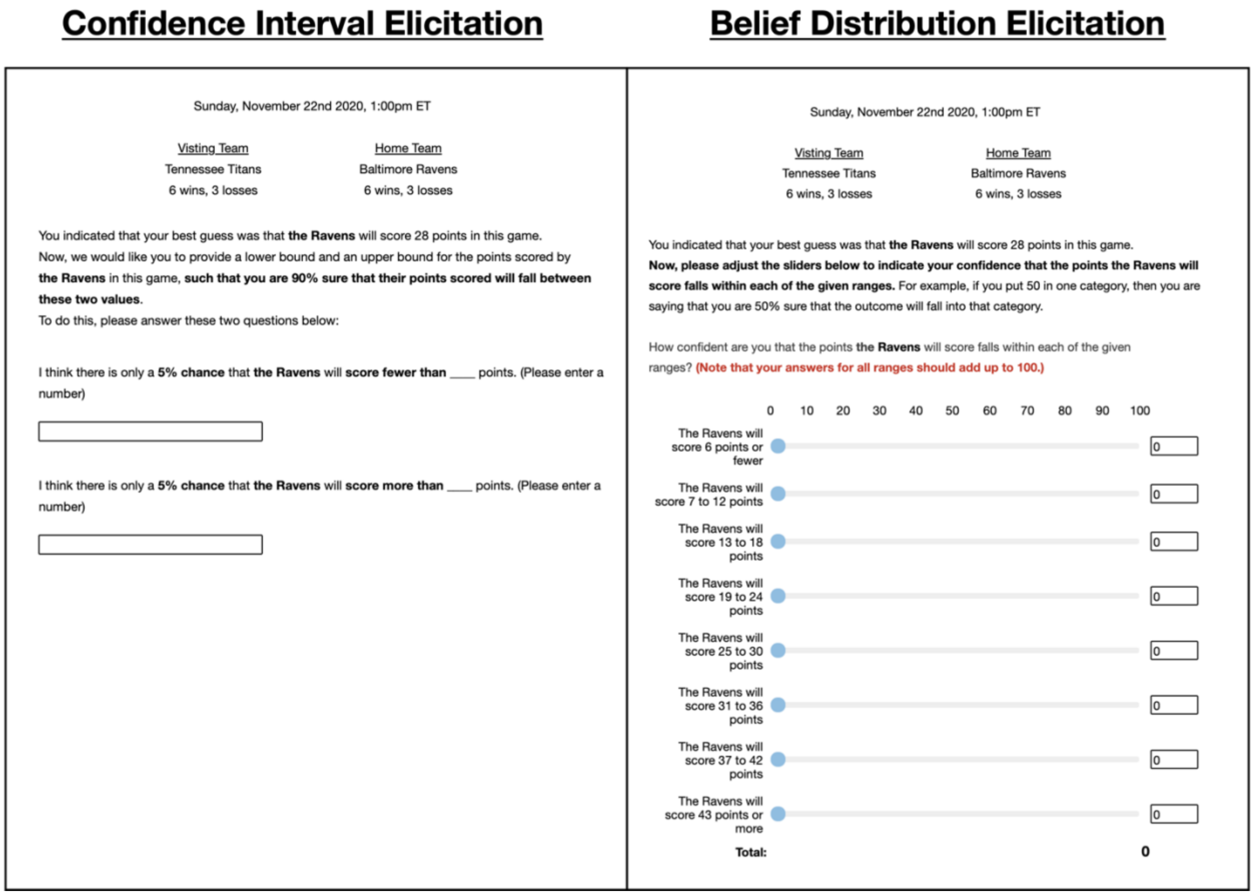

Figure 3. Example of Confidence Interval and Belief Distribution Elicitations in Study 2

Note. Screenshots of the Confidence Interval and the Belief Distribution conditions for one game in Study 2. The bolded and colored terms are from the original survey presented to participants.

At the end of both studies (and of Study 6 in this paper), for exploratory purposes we asked participants five questions designed to assess participants' knowledge of the NFL. Specifically, they were asked to identify which NFL team five players currently play for and were asked to respond to these questions without looking up the answers. We did not observe significant interactions between the NFL knowledge score and our experimental conditions, so we do not discuss this measure further.

Results and Discussion

Preliminary analyses. Before presenting the main results, we would like to do two things. First, we'd like to establish that participants in our studies were *over*confident. Obviously, we cannot compare participants' 9-point scale ratings of confidence to some objective benchmark, and so cannot properly assess whether participants are overconfident on this measure. However, we can assess whether participants' likelihood estimates were too high/overconfident in these studies. Specifically, we can compare how likely they said their predictions were to be accurate to how accurate these predictions actually were. In Study 1, participants' predictions were accurate 59.5% of the time, but their average likelihood estimate was significantly higher (68.9%), $b = 11.50$, $SE = 2.27$, $t = 5.06$, $p < .001$. In Study 2, participants' predictions were within 5 points of the correct answer only 37.7% of the time, but their average likelihood estimate was 60.5%, $b = 23.37$, $SE = 2.68$, $t = 8.73$, $p < .001$. Indeed, as shown in Table S3.1 of Supplement 4, participants' likelihood estimates were directionally overconfident in every condition of every study, and often by a very large margin. This means that, for all of our studies, whenever an intervention increased confidence, it also increased *over*confidence.

Second, as mentioned earlier, past research investigated the effect of providing a belief distribution on overprecision by comparing the width of the confidence intervals across conditions (e.g., Haran et al., 2010). As discussed above, this measure is potentially problematic, because participants who provide an entire belief distribution may feel compelled to allocate some probability to outcomes that they believe to be impossible, which could artificially widen their confidence intervals. While not the primary aim of our investigation, in many of our studies (Studies 1-3, 6-7, and S3) we were able to compare the width of the 90% confidence interval between the Belief Distribution condition and the Confidence Interval condition.[5] As shown in

[5] In the Belief Distribution condition, we computed the width of the 90% confidence interval using the algorithm developed by Haran et al. (2010). The algorithm requires the range of outcomes to be bounded by a minimum and a

Table S4 of Supplement 5, in most of our studies we did replicate past findings: Providing a belief distribution led to significantly wider confidence intervals than merely stating the 90% confidence interval.[6] As you will see, this result is inconsistent with what we find on our other – and arguably less problematic – measures of overconfidence.

Main analyses. If providing a belief distribution truly reduces (over)confidence, then participants in the Belief Distribution condition should have provided a lower confidence rating, a lower likelihood estimate, and been less likely to wager on their predictions, compared with participants in the Confidence Interval and Control conditions. However, this is not what we found. As shown in Figure 4, providing a belief distribution did not reduce confidence on any measure; on the contrary, it sometimes significantly increased confidence.

In both studies, we regressed participants' confidence on their experimental condition(s) (contrast coded in Study 1; dummy coded in Study 2), while including fixed effects for the predicted game (Study 1) or the predicted team (Study 2) and clustering standard errors by participant. Relative to providing a 90% confidence interval, providing an entire belief distribution had no influence on the confidence measures in Study 1 (b = .13, SE = .10, t = 1.34, p = .180 for confidence ratings; b = -.90, SE = .91, t = -1.00, p = .319 for likelihood estimates) and directionally increased confidence in Study 2 (b = .21, SE = .14, t = 1.52, p = .130 for confidence ratings; b = 3.09, SE = 2.05, t = 1.51, p = .132 for likelihood estimates; b = .09, SE = .05, t = 1.89, p = .059 for willingness to wager). Relative to providing no belief distribution or confidence interval (the Control condition in Study 2), providing a belief distribution significantly increased participants'

maximum. We used -40 and 40 in Study 1's calculation and 0 and 48 in Study 2's calculation. To keep the upper bound and lower bound consistent, we winsorized the values in the Confidence Interval condition to the same minimum and maximum. In addition, for "backwards" confidence intervals – ones with higher lower bounds than upper bounds – we pre-registered to treat them as equal to 0 in Study 1 and as equal to the absolute difference between the two values in Study 2. Results excluding responses with backwards confidence intervals do not differ meaningfully.

[6] Because we wanted to ensure that we were able to replicate past research, we did pre-register to conduct this analysis in some (but not all) of our studies, including Studies 1 and 2. See Supplement 5 for details.

confidence in their predictions, $b = .58$, $SE = .14$, $t = 4.12$, $p < .001$, their likelihood estimates, $b = 4.31$, $SE = 2.07$, $t = 2.08$, $p = .038$, and directionally increased their likelihood of wagering on their predictions, $b = .07$, $SE = .05$, $t = 1.57$, $p = .117$.

In sum, in Studies 1 and 2 we found no evidence that providing a belief distribution reduces confidence in one's predictions. Instead, we sometimes found the opposite, that providing a belief distribution sometimes makes people significantly *more* confident in their initial predictions.

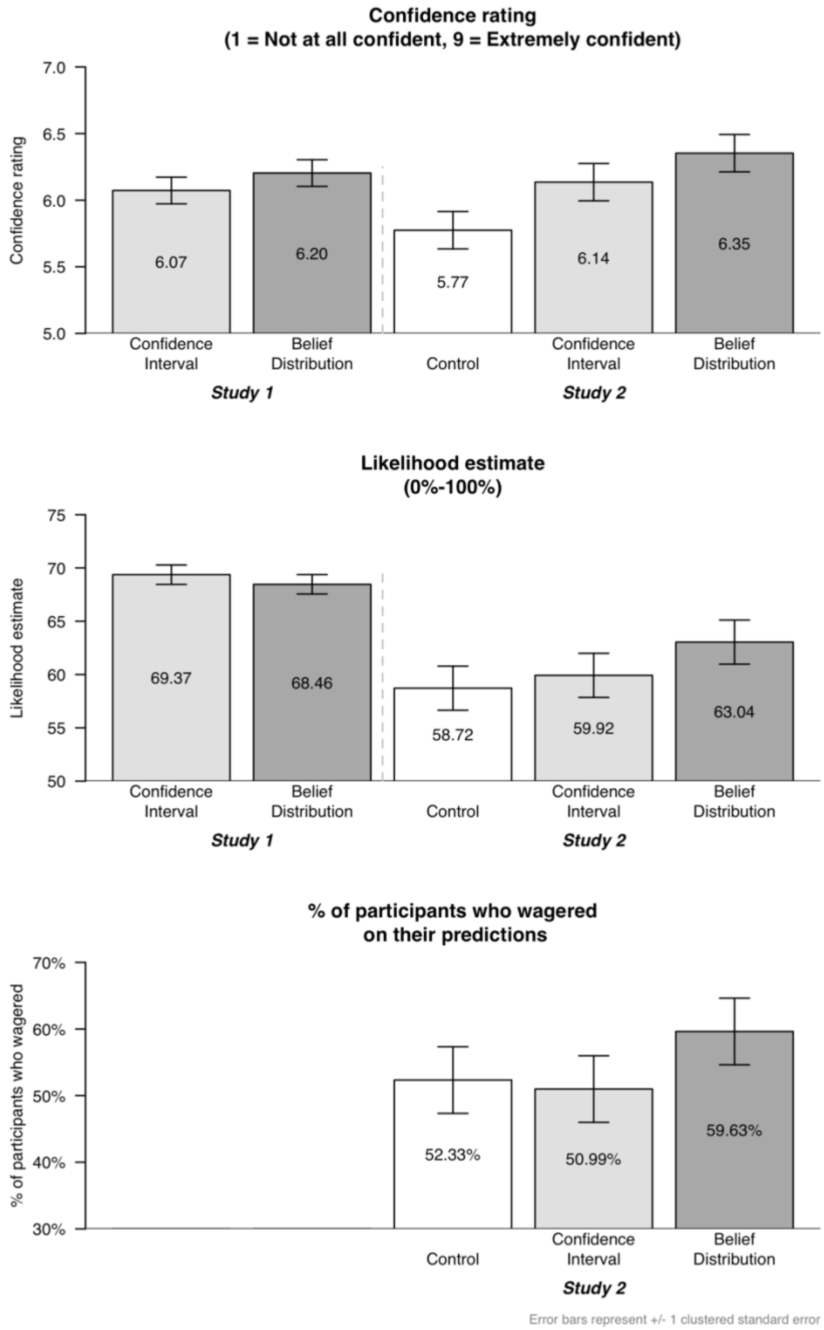

Figure 4. Studies 1 and 2: Means and Percentages

Note. Error bars represent +/- 1 clustered standard error.

Study 3

In Studies 1 and 2, participants in the Belief Distribution condition always made predictions *before* constructing the belief distribution. This may have led them to construct belief distributions that served to rationalize and reinforce their initial predictions. If that is the case, then perhaps providing belief distributions would reduce overconfidence if they were elicited *before* providing a specific prediction, since people would presumably feel less pressure to provide a distribution that rationalizes a prediction they have not yet made. To test this, in Study 3 we randomly assigned participants to provide either a belief distribution or confidence interval before or after they made a specific prediction, and then assessed how confident they were in that prediction.

Method

Participants. We conducted Study 3 using U.S. participants from Prolific. We decided in advance to recruit 1,700 participants. Only participants who passed the attention check at the beginning of the survey were allowed to proceed to the survey. We pre-registered to exclude all responses from duplicate Prolific IDs or duplicate IP addresses (147 exclusions) and participants who misreported their Prolific IDs (12 exclusions). We wound up with a final sample of 1,816 participants.[7] The sample averaged 33.9 years of age and was 54.7% female.

Procedure. To try to better establish the generalizability of the results that emerged in Studies 1 and 2, in Study 3 we moved beyond sports predictions and instead asked participants to predict the percentage of all survey respondents who would express certain preferences.

At the beginning of the study, participants answered four binary preference questions about themselves (e.g., "Do you prefer Thanksgiving or Christmas?"). They were then asked to estimate for each question the percentage of survey respondents who would prefer a particular option (e.g., "My best guess is that _____% of survey respondents prefer Thanksgiving to Christmas."). In all

[7] We don't know why we ended up with more participants than we requested. We assume it was a glitch with Prolific.

studies using this prediction context, we randomized the target option for prediction (e.g., either "prefer Thanksgiving" or "prefer Christmas") and the order in which the prediction questions were presented. Table 2 presents the exact wording and the true percentages for all the preference questions we used in Studies 3-5 and Studies 7-8. Due to the change of prediction domain, the range of possible outcomes presented in the Belief Distribution condition was necessarily different. In Studies 3-5 and 7-8, participants in the Belief Distribution condition allocated probabilities to 10 categories covering the entire range of outcomes: 0-10%, 11-20%, …, 91-100%.

As in Study 2, we randomly assigned participants to provide a belief distribution (Belief Distribution condition), a 90% confidence interval (Confidence Interval condition), or neither of these (Control condition). For those in the Belief Distribution condition and the Confidence Interval condition, we additionally manipulated whether they provided their best estimate before or after providing the belief distribution or the confidence interval. Participants who were assigned to make their predictions first (Best Estimate First condition) followed the same procedure as in Studies 1 and 2, first providing their best estimate, and then, on a second page, answering the confidence interval questions or providing their belief distribution. Participants assigned to make their predictions afterward (Best Estimate Last condition) first provided their confidence interval or belief distribution, and then, on a second page, provided their best estimate. Thus, participants were randomly assigned to one of five conditions in this study: Control vs. Belief Distribution/Best Estimate First vs. Confidence Interval/Best Estimate First vs. Belief Distribution/Best Estimate Last vs. Confidence Interval/Best Estimate Last.

Table 2. Studies 3-5, 7 and 8: Wording and True Percentages for Preference/Behavior Questions

Preference/Behavior questions	% of participants who chose the first option
Study 3	
Do you prefer Thanksgiving or Christmas?	24%
Would you prefer to be able to see the future or change the past?	52%
Would you prefer to have 1 wish granted today or 3 wishes granted in 5 years?	40%
Would you prefer to have more money or more time?	73%
Study 4	
Do you prefer pasta or pizza?	31%
Do you prefer a vacation in the mountains or at the beach?	43%
Do you prefer spending money or saving money?	28%
Would you prefer to have photographic memory or an extra gain of 40 IQ points?	60%
Study 5 (Moderate condition)	
Do you prefer milk chocolate or dark chocolate?	60%
Which ice cream flavor do you prefer: chocolate or vanilla?	52%
Do you have an iPad? (Yes/No)	48%
Study 5 (Extreme condition)	
Do you prefer milk chocolate or wasabi-flavored chocolate?	99%
Which ice cream flavor do you prefer: chocolate or cheese?	95%
Do you have a TV? (Yes/No)	97%
Study 7	
Do you prefer pancakes or waffles?	49%
Which superpower would you prefer to have: invisibiltiy or time travel?	34%
If you were going to take a walk, would you prefer listening to music or listening to a podcast?	73%
Do you prefer the smell of freshly brewed coffee or the smell of freshly baked cookies?	49%
Study 8	
Would you prefer to type amazingly fast or to read amazingly fast?	31%
Are you a morning person or a night person?	40%
Would you prefer to have more money or more fame?	98%
What do you think is worse: doing laundry or doing dishes?	35%

Results and Discussion

We pre-registered to conduct two sets of analyses. In our first set of analyses, we ignored the Control condition and regressed the dependent measures on (1) the Belief Distribution/Confidence Interval condition (contrast-coded), (2) the Best Estimate First/Last condition (contrast-coded), and (3) their interaction. This allowed us to examine whether task order moderated the effect of providing a belief distribution on (over)confidence. In our second set of analyses, we compared the Control condition to all the other conditions by regressing the dependent measures on indicators for the Belief Distribution conditions and the Confidence Interval conditions. We included fixed effects for prediction items and clustered standard errors by participant in both sets of analyses.

We present the results in Figure 5. First, we found that compared to providing a confidence interval (Confidence Interval condition), providing a belief distribution (Belief Distribution condition) significantly increased participants' confidence in their predictions ($b = .24$, $SE = .08$, $t = 2.89$, $p = .004$ for confidence ratings, and $b = 3.37$, $SE = 1.17$, $t = 2.87$, $p = .004$ for likelihood estimates). Importantly, this effect did not depend on whether participants' confidence intervals or belief distributions were elicited before or after they made their specific predictions, as the interaction was not significant for either the confidence ratings ($p = .598$) or the likelihood estimates ($p = .474$). Moreover, comparing the Control condition to the two Belief Distribution conditions, we found that providing a belief distribution significantly increased confidence, regardless of whether participants made the predictions before or after giving the belief distribution (confidence rating: $b = .34$, $SE = .11$, $t = 3.04$, $p = .002$ and $b = .23$, $SE = .12$, $t = 1.98$, $p = .048$ for making the predictions before and after providing the belief distribution, respectively; likelihood estimate: $b = 4.18$, $SE = 1.62$, $t = 2.58$, $p = .010$ and $b = 4.64$, $SE = 1.64$, $t = 2.83$, $p = .005$ for making the predictions before and after providing the belief distribution, respectively). The two Confidence Interval conditions and the Control condition did not differ significantly on either dependent measure (ps $\geq .299$). These results suggest that providing belief distributions may increase rather than decrease confidence, even when participants construct those distributions before their specific predictions are elicited. The increase in confidence, however, did not come with an increase in accuracy regardless of the order of the task and was thus unwarranted.[8]

[8] In Study 3, we also pre-registered to analyze the absolute error of participants' estimates (i.e., the absolute difference between their best estimates and the truth). We wanted to analyze this measure primarily to see whether participants who provided belief distributions or confidence intervals prior to making their predictions were more or less accurate than those who provided belief distributions or confidence intervals after making their predictions. In an analysis

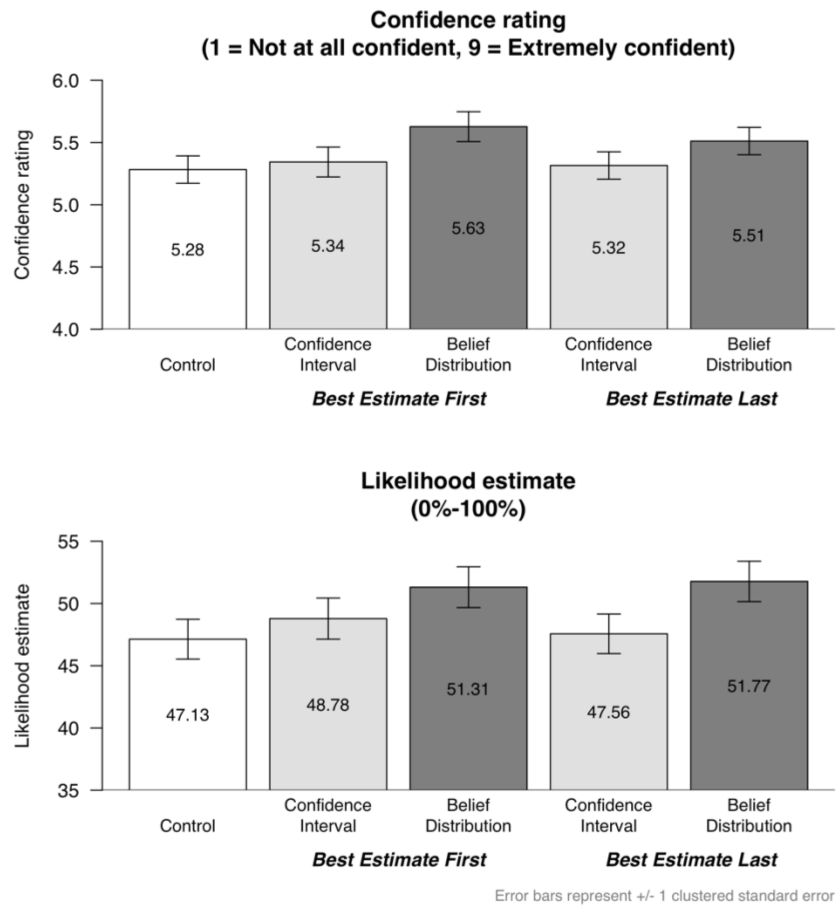

Figure 5. Study 3 Results

Note. Error bars represent +/-1 clustered standard error.

Study 4

comparing the four other conditions to the Control condition, we found that those in the Confidence Interval/Best Estimate First condition provided marginally more accurate predictions (b = -1.09, SE = 0.56, t = -1.96, p = .050). Because this condition and the Control condition were procedurally identical up to the point at which predictions were provided, this simply reflects a small failure of random assignment. None of the other conditions had significantly more or less error than the Control condition (ps > .428). We also conducted a regression that omitted the Control condition and analyzed the other four conditions as a 2 (Belief Distribution vs. Confidence Interval) x 2 (Best Estimate First vs. Last). We found only a barely significant main effect of timing, indicating that predictions were more accurate when they were made before the belief distribution or confidence interval elicitations (b = -0.76, SE = 0.38, t = -1.98, p = .047). The other effects were nonsignificant (ps > .326). Altogether, this means that providing a belief distribution or confidence interval before making predictions did not increase the accuracy of those predictions.

In the first three studies, participants reported how confident they were that their prediction was within a correct range. But perhaps providing belief distributions is more likely to reduce (over)confidence when people are asked to consider how confident they are that their answer is exactly right rather than within some range, because constructing a belief distribution may make it salient that it is hard to get an answer exactly right. To test this, in Study 4, we investigated whether the effect of providing a belief distribution on overconfidence is moderated by whether people are asked to make an imprecise prediction (e.g., "How confident are you that your prediction is within 5 percentage points of the correct answer?") or a precise prediction (e.g., "How confident are you that your prediction is correct?"). We randomly assigned participants to one cell of a 2 (Belief Distribution vs. Control condition) x 2 (Low Precision vs. High Precision condition) between-subjects design. In the Low Precision condition, we asked them to estimate the percentage of survey respondents who held particular preferences (as in Study 3). In the High Precision condition, we asked them to estimate exactly how many of 10 randomly selected respondents held particular preferences.

Method

Participants. We conducted Study 4 using U.S. participants from Amazon's Mechanical Turk (MTurk). We decided in advance to recruit 1,300 participants. Only participants who passed the attention check at the beginning of the survey were allowed to proceed to the survey. We pre-registered to exclude all responses from duplicate MTurk IDs or duplicate IP addresses (81 exclusions) and participants who misreported their MTurk ID (6 exclusions). We wound up with a final sample of 1,213 participants. The sample averaged 39.1 years of age and was 44.9% female.

Procedure. As in Study 3, participants began the study by answering four binary preference questions about themselves (e.g., "Do you prefer pizza or pasta?"; see Table 2). They were then randomly assigned to one cell of a 2 (Belief Distribution vs. Control condition) x 2 (Low Precision vs. High Precision condition) between-subjects design. Those in the Low Precision condition were asked to estimate for each question the percentage of survey respondents who would prefer a particular option (e.g., “My best guess is that _____% of survey respondents prefer pizza to pasta.”). Then, after providing a belief distribution or not, they were asked to rate how confident they were and how likely it was that their prediction was within 5 percentage points of the correct answer. Those in the High Precision condition read that “We randomly selected 10 survey respondents” and they were asked to predict how many of those 10 respondents would prefer a particular option (e.g., “My best guess is that _____ of the 10 randomly selected survey respondents prefer pizza to pasta.”). Then after providing a belief distribution or not, they were asked to rate how confident they were and how likely it was that their prediction was correct (see Table 1 for exact wordings).

Note that we designed this precision manipulation in a way that was intended to hold constant the difficulty of the prediction task. In other words, we expected that making a prediction that was within 5 percentage points of the correct answer on a 101-percentage-point scale (as in the Low Precision condition) to be no more or less difficult than correctly predicting the correct answer on a 0-10 scale (as in the High Precision condition). And, indeed, participants in the Low Precision condition were not more or less likely to make a prediction that was within 5 percentage points of the correct answer (M = 16.7%) than were participants in the High Precision condition to make a prediction that was equal to the correct answer (M = 16.5%), b = .002, SE = .01, p = .860.

Results and Discussion

We regressed the dependent measures on (1) the Belief Distribution condition (contrast-coded), (2) the High Precision condition (contrast-coded), and (3) their interaction. We included fixed effects for prediction items and clustered standard errors by participants.

We present the results in Figure 6. As you can see, providing a belief distribution had no significant influence on participants' confidence in their predictions (b = .04, SE = .09, t = .39, p = .697 for confidence ratings, and b = -.95, SE = 1.30, t = -.73, p = .468 for likelihood estimates), and this result was not moderated by the precision condition (the interaction was p = .989 for confidence ratings and p = .433 for likelihood estimates). Though not of primary interest, we did find that participants were more confident in their predictions in the High Precision condition than in the Low Precision condition (b = .17, SE = .09, t = 1.82, p = .069 for confidence ratings, and b = 4.05, SE = 1.30, t = 3.11, p = .002 for likelihood estimates). Overall, these results suggest that providing a belief distribution does not reduce people's confidence in their predictions, even when these predictions are precise.

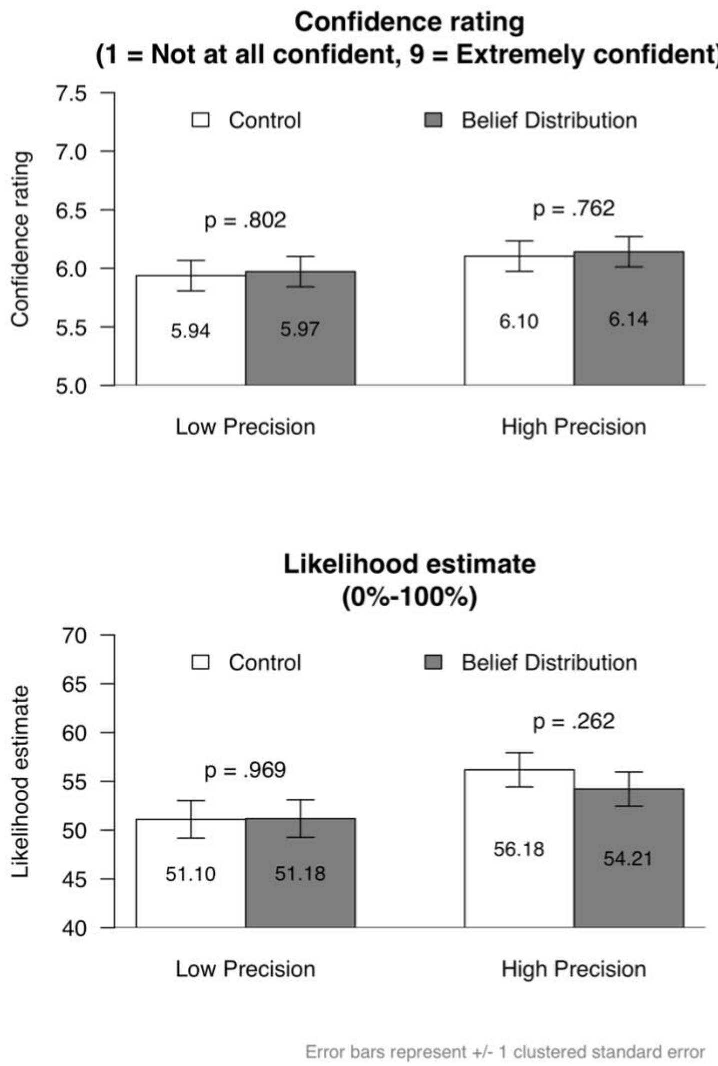

Figure 6. Study 4 Results.

Note. Error bars represent +/-1 clustered standard error.

Study 5

In Study 5, we sought to investigate whether the effect of providing a belief distribution on overconfidence would be moderated by whether the correct answer was extreme and somewhat obvious or moderate and somewhat uncertain. Our thinking was as follows. When a correct answer is moderate, there might be a fair bit of uncertainty associated with one's prediction. Providing a belief distribution might make that uncertainty salient, thereby reducing confidence in that prediction. Conversely, when a correct answer is extreme, there might be a great deal of confidence associated with that answer, and that confidence might be reinforced when people are asked to provide a belief distribution.

Method

Participants. We conducted Study 5 using U.S. participants from MTurk. We decided in advance to recruit 1,300 participants. Only participants who passed the attention check at the beginning of the survey were allowed to proceed to the survey. We pre-registered to exclude all responses from duplicate MTurk IDs or duplicate IP addresses (6 exclusions) and participants who misreported their MTurk IDs (22 exclusions). We wound up with a final sample of 1,277 participants. The sample averaged 40.2 years of age and was 55.1% female.

Procedure. This study was identical to Study 3, except that (1) we asked participants three questions instead of four, (2) those questions were different (see Table 2), and (3) within question we manipulated whether the answer was Extreme or Moderate. For example, the Extreme version of one question read, “Which ice cream flavor do you prefer: chocolate or cheese?” whereas the Moderate version read, “Which ice cream flavor do you prefer: chocolate or vanilla?” In a pretest (see Study S2 in the Supplement 9), the choice shares for the Extreme versions were either above 90% or below 10% (e.g., the vast majority of participants chose chocolate ice cream over cheese ice cream), whereas the choice shares for the Moderate versions were between 30% and 70% (e.g., roughly half of the participants chose chocolate ice cream over vanilla ice cream).

Results and Discussion

First, it is worth noting that the true percentages of extreme and moderate versions of the questions suggest that our manipulation was successful: The percentage of participants choosing the most popular option ranged from 95% to 99% for the extreme version and from 52% to 60% for the moderate version (see Table 2). In addition, participants' best estimates suggest that they intuited those differences: The average of participants' forecasts ranged from 83% to 89% for the extreme questions and from 51% to 61% for the moderate questions.

For our main analysis, we regressed the dependent measures on (1) the Belief Distribution condition (contrast-coded), (2) the Extreme condition (contrast-coded), and (3) their interaction. We included fixed effects for prediction items and clustered standard errors by participants.

We present the results in Figure 7. Replicating some of our previous studies, we found that asking participants to provide a belief distribution significantly *increased* their confidence in their predictions (b = .33, SE = .08, t = 4.10, p < .001 for confidence ratings and b = 3.95, SE = 1.25, t = 3.16, p = .002 for likelihood estimates). Importantly, the extremity of the true answers did not moderate this effect (the interaction was p = .820 for confidence ratings and p = .990 for likelihood estimates). We also (quite sensibly) found that people were more confident in their predictions for the extreme questions than for the moderate questions (b = 1.64, SE = .06, t = 26.29, p < .001 for confidence ratings and b = 23.25, SE = .93, t = 24.94, p < .001 for likelihood estimates). Overall, these results suggest that providing a belief distribution is more likely to increase than decrease people's confidence, and this is true regardless of whether the correct answers are extreme or moderate.[9]

[9] In this study, we also pre-registered to analyze whether participants in the Belief Distribution condition were more likely to assign higher probabilities to the category that included their best estimate for items with Extreme answers than for items with Moderate answers. And, indeed, this is what we found: b = 26.41, SE = 1.59, t = 16.66, p < .001. Thus, participants were in fact more likely to give more dispersed (and potentially less belief-reinforcing) belief distributions when the correct answers were moderate than when they were extreme.

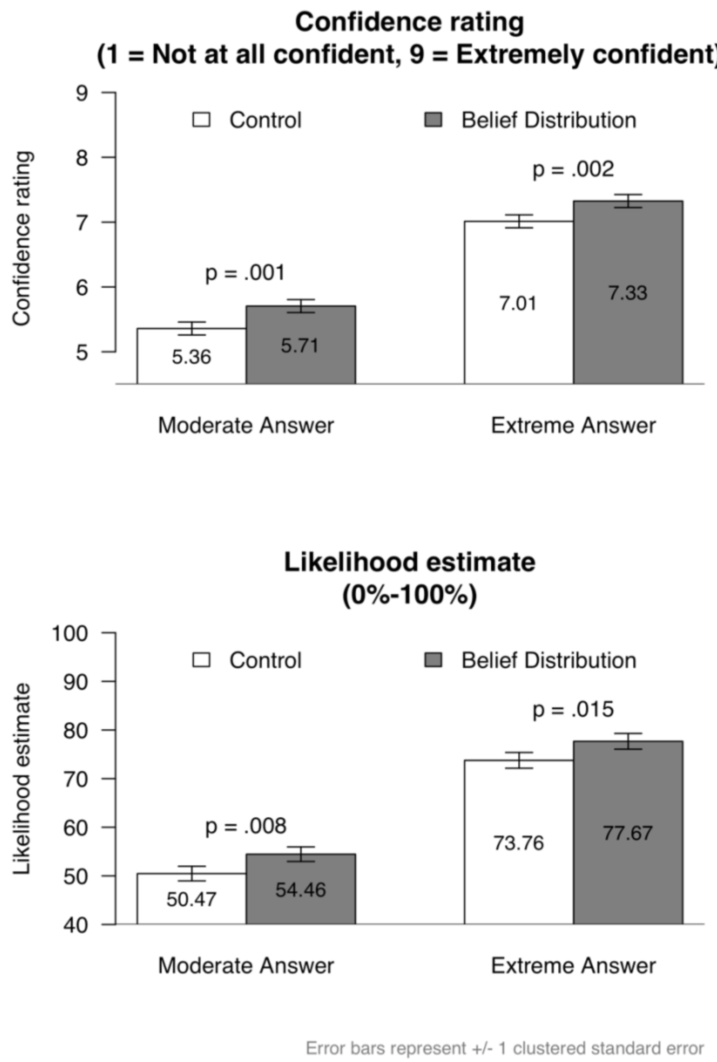

Figure 7. Study 5 Results.

Note. Error bars represent +/-1 clustered standard error.

Studies 6-8

In Studies 1-5, we found that providing a belief distribution sometimes significantly increased people's confidence, and never significantly decreased it. These results are surprising because providing a belief distribution should prompt people to consider more possibilities, which should act to reduce overconfidence (Haran et al., 2010; Moore, 2020, 2022). Might other interventions that capitalize on this principle work better to decrease confidence? If so, how they differ from constructing belief distributions may shed light on why constructing belief distributions sometimes increases confidence.

In Studies 6-8, we designed and tested some interventions that we thought would be more likely to work to decrease overconfidence, all with the underlying goal of encouraging people to think about ways in which their original estimate might be incorrect. In Studies 6 and 7, we tested a Multiple Guesses intervention, in which participants were asked to provide multiple estimates for the same prediction. We thought that asking participants to provide multiple predictions might make them realize that many different outcomes were likely, thus reducing their confidence in their initial prediction. In Study 8, we tried two additional interventions, a Surprise intervention that asked participants to indicate how surprised they would be if the outcome fell within each of a set of mutually exclusive and collectively exhaustive ranges, and a Choosing Possibilities intervention that asked participants to simply indicate which outcomes were at all possible (without allocating probabilities to each outcome). Like the belief distribution interface, both interventions also showed the entire range of possible outcomes. We thought that the Surprise intervention might reduce confidence by cuing participants to the notion that there are many different outcomes that would not be terribly surprising, and that the Choosing Possibilities intervention might reduce confidence by making salient that many different outcomes could transpire.

Method

Participants. We conducted Studies 6 and 7 using U.S. participants from Prolific and Study 8 using U.S. participants from MTurk. We decided in advance to recruit 1,300 participants for all three studies. We pre-registered to exclude all responses from duplicate Prolific/MTurk IDs or duplicate IP addresses (64 exclusions, 53 exclusions, and 9 exclusions, respectively), participants who misreported Prolific/MTurk IDs (3 exclusions, 5 exclusions, and 17 exclusions, respectively) and participants who failed the attention check questions (41 exclusions, 18 exclusions, and 10 exclusions, respectively). We wound up with final samples of 1,186, 1,242, and 1,275 participants, respectively. The samples averaged 35-40 years of age and were 26% female for Study 6 (an NFL study) and 54% and 57% female for Studies 7 and 8, respectively.

Procedures. Studies 6-8 followed similar procedures as in previous studies, so here we simply describe the ways in which they were different.

Study 6. In Study 6, participants predicted the total points scored by a team for four upcoming NFL games played on November 1, 2020. The study's procedure was exactly the same as that of Study 2, except for two changes. First, in addition to the Control condition, the Confidence Interval condition, and the Belief Distribution condition, we added a Multiple Guesses condition. In this condition, participants were asked to give their best, second best, and third best guesses for the points scored by a randomly selected team. In this condition, we asked participants to indicate how confident they were in their (first) best guess. Similar to how constructing belief distributions should encourage people to consider different outcomes, we expected that simply asking participants to give multiple guesses should bring to mind a wider range of outcomes and thus reduce confidence in the first best guess. Second, in this study we slightly altered the dependent measures, so that we assessed how confident participants would be about their prediction being within 3 points (rather than 5 points) of the correct answer. See Table 1 for exact question wordings.

Study 7. Study 7 examined whether the results in Study 6 would replicate in the preference prediction domain. The conditions and the procedures in each condition were identical to Study 6, except for necessary changes to the prediction items (see Table 2) and to the exact wording of the dependent measures (see Table 1).

Study 8. In Study 8, participants were asked to predict the percentage of participants who would choose a certain option in response to four preference questions (see Table 2). Participants were randomly assigned to one of four conditions: Control vs. Belief Distribution vs. Surprise vs. Choosing Possibilities. The procedures in the Control and the Belief Distribution conditions were identical to those in previous studies. In the Surprise condition, participants were presented with the same 10 ranges of possible outcomes as in the Belief Distribution condition (0-10%, 11-20%, …, 91-100%). But instead of allocating probabilities, they were asked to indicate for each range of outcomes how surprised they would be if the correct answer fell within that range (1 = not at all surprised, 7 = extremely surprised). In the Choosing Possibilities condition, participants also saw the same 10 ranges of possible outcomes (0-10%, 11-20%, …, 91-100%). But they were asked to simply select all the ranges that they thought might possibly contain the correct answer. They did not allocate probabilities to each range.

It is worth emphasizing that both the Surprise condition and the Choosing Possibilities condition showed participants the same range of possible outcomes as the Belief Distribution condition. The key difference is that participants were not asked to allocate probabilities to possible outcomes. Therefore, if either condition did not increase confidence as much as the Belief Distribution condition, then the aspects in which their designs deviate from the Belief Distribution condition might suggest possible mechanisms through which constructing a belief distribution increases confidence.

Results and Discussion

In each study, we ran three sets of OLS regressions. We omitted one condition in each set of analyses and regressed the dependent measures on the three dummies for the other three conditions. This allowed us to examine all pairwise comparisons among all conditions. In each regression, we included fixed effects for the team (Study 6) or the prediction item (Studies 7 and 8) and clustered standard errors by participants.

We present the results in Table 3. First, it is worth noting that we replicated the results of our previous studies. Compared to the Control condition, asking participants to provide a belief distribution significantly increased confidence ratings in Studies 6 (b = .54, SE = .12, t = 4.30, p < .001) and 8 (b = .26, SE = .12, t = 2.24, p = .025), marginally increased confidence ratings in Study 7 (b = .25, SE = .14, t = 1.78, p = .075), and directionally increased likelihood estimates in all three studies (b = 1.67, SE = 1.69, t = .99, p = .322 in Study 6; b = 2.80, SE = 1.77, t = 1.58, p = .114 in Study 7; b = 2.33, SE = 1.75, t = 1.33, p = .184 in Study 8).

Table 3. Studies 6-8 Results

	Control		Confidence Interval		Belief Distribution		Multiple Guesses		Surprise		Choosing Possibilities	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Confidence rating (1 = Not at all confident, 9 = Extremely confident)												
Study 6	5.02[a]	1.84	5.27[a,b]	1.82	**5.55[c]**	1.76	**5.31[b]**	1.75				
Study 7	5.18[a]	2.06	5.37[a,b]	1.96	5.43[a,b]	1.99	**5.49[b]**	1.92				
Study 8	5.45[a]	1.87			**5.72[c]**	1.93			5.42[a,b]	1.95	5.23[b]	1.96
Likelihood estimate (0% - 100%)												
Study 6	49.28[a,b]	23.50	47.35[b]	25.51	50.95[a]	23.74	49.43[a,b]	22.95				
Study 7	47.23[a]	26.32	48.89[a,b]	26.13	50.03[a,b]	26.09	**51.24[b]**	26.25				
Study 8	49.77[a,b]	27.04			52.10[b]	27.33			52.61[b]	25.37	47.96[a]	26.85

Note. Within each row, **boldface** indicates that participants in this condition were significantly *more* confident than the Control condition ($p < .05$), underlining indicates that participants in this condition were significantly *less* confident than the Control condition ($p < .05$). Means sharing the same superscript are not significantly different from each other ($p < .05$).

Did any of the new interventions successfully reduce people's confidence? First, let's consider the Multiple Guesses condition, which we included in both Studies 6 and 7. Contrary to our expectation, we found that asking participants to provide multiple guesses tended to *increase* participants' confidence (confidence ratings: b = .30, SE = .12, t = 2.38, p = .017 in Study 6 and b = .31, SE = .13, t = 2.27, p = .023 in Study 7; likelihood estimates: b = .13, SE = 1.65, t = .08, p = .936 in Study 6 and b = 4.00, SE = 1.79, t = 2.23, p = .026 in Study 7). So that intervention was certainly not successful at reducing overconfidence.

Relative to the Control condition, asking participants to indicate how surprised they would be if the outcome were to fall within each possible range (i.e., the Surprise condition of Study 8) did not increase overconfidence, but it did not significantly decrease it either (b = -.03, SE = .11, t = -.27, p = .786 for confidence ratings, and b = 2.84, SE = 1.66, t = 1.71, p = .087 for likelihood estimates).

Of the three interventions that we tested, the Choosing Possibilities one was most promising. Relative to the Control condition, asking participants to merely choose which ranges might possibly contain the right answer (barely) significantly reduced their confidence ratings (b = -.23, SE = .12, t = -1.99, p = .047) and directionally reduced their likelihood estimates (b = -1.81, SE = 1.69, t = -1.07, p = .286). That this intervention produced such different results than the Belief Distribution was quite interesting to us, as the only procedural difference between these conditions was that the Belief Distribution condition asked participants to allocate probabilities to each range of outcomes, while the Choosing Possibilities condition asked them merely to indicate which ranges of outcomes might possibly include the correct answer. This suggests that overconfidence may not be increased by being asked to consider the full range of outcomes, but more specifically by the act of allocating probabilities to that range of outcomes. We conducted Studies 9 and 10 to test this idea more directly.

Studies 9-10

The results of Study 8 led us to suspect that merely seeing and thinking about all possible outcomes does not increase overconfidence; rather, overconfidence may be specifically increased by asking people to allocate probabilities to a wide range of outcomes. Why would this be? Perhaps it arises because people tend to allocate probabilities in a way that suggests that most outcomes are in fact unlikely, and/or that their own forecasted outcome is especially likely. In other words, perhaps they allocate probabilities in a way that reinforces their pre-existing beliefs.

In Studies 9 and 10, we asked participants to predict National Basketball Association (NBA) game outcomes, and we randomly assigned them to one of four conditions. In addition to the Control condition, the Belief Distribution condition, and the Choosing Possibilities condition, we added a new Choosing Possibilities + Belief Distribution condition. Participants in this new condition first chose which ranges of outcomes might contain the correct answer and then provided a belief distribution.

If thinking about all possible outcomes does not increase overconfidence, but allocating probabilities to those outcomes does increase overconfidence, then we should find that this new Choosing Possibilities + Belief Distribution manipulation increases overconfidence, both relative to the Control condition, and relative to the Choosing Possibilities condition.

Method

Participants. We conducted Studies 9 and 10 using U.S. participants from Prolific. We decided in advance to recruit as many participants as we could by 8 pm Eastern Time on the game day (the starting time of the earliest game that we asked participants to forecast), so long as we did not exceed 1,300 participants. We pre-registered to exclude all responses from duplicate Prolific IDs or duplicate IP addresses (95 exclusions and 60 exclusions, respectively), participants who misreported Prolific IDs (11 exclusions and 4 exclusions, respectively), and participants who failed the attention check question (87 exclusions and 75 exclusions, respectively). We wound up with final samples of 1,105 and 946 participants, respectively. The samples averaged 30-31 years of age and were 44%-45% female.

Procedures. In each study, participants were asked to predict the outcomes of four upcoming NBA games. The procedures were similar to those in previous NFL studies. For each game, participants predicted the total points scored by a randomly selected team. We assessed confidence by asking participants to rate how confident they were that their prediction would be within 5 points of the correct answer, and to indicate how likely their prediction was to be within 5 points of the correct answer (see Table 1).

In both studies, participants were randomly assigned to one of four conditions: Control vs. Choosing Possibilities vs. Belief Distribution vs. Choosing Possibilities + Belief Distribution. The Control, Belief Distribution, and Choosing Possibilities conditions followed the same procedures as in previous studies. In the Belief Distribution condition and the Choosing Possibilities condition, participants saw nine categories covering the entire range of possible outcomes (labeled as “Below 80,” “80-89,” “90-99,” “100-109,” “110-119,” “120-129,” “130-139,” “140-149,” “150 and above”). In the new Choosing Possibilities + Belief Distribution condition, participants first completed the Choosing Possibilities task, and then on a subsequent page they completed the Belief Distribution task.

Studies 9 and 10 were identical in procedures except for two differences. First, the set of games were necessarily different. The games in Study 9 were played on February 12th, 2021, and the games in Study 10 were played on February 19th, 2021. Second, we reversed the elicitation order in the two studies. In Study 9, participants who were not in the Control condition gave their best estimate prediction *before* completing either the Choosing Possibilities and/or the Belief Distribution task. In Study 10, participants who were not in the Control condition gave their best estimate prediction *after* completing the Choosing Possibilities and/or the Belief Distribution task.

At the end of the survey, we asked participants five NBA knowledge questions, in which they were asked to identify which NBA team five players currently play for. These questions were exploratory.

Results and Discussion

In each study, we ran three sets of OLS regressions. We omitted one condition in each set of analyses and regressed the dependent measures on dummy-coded indicators for each of the other three conditions. This allowed us to test all pairwise comparisons between all conditions. We included fixed effects for the predicted team and clustered standard errors by participants.

We present the results in Figure 8. Replicating previous results, we found that providing a belief distribution (Belief Distribution condition) significantly increased confidence compared to the Control condition (confidence rating: b = .44, SE = .13, t = 3.37, p = .001 in Study 9 and b = .48, SE = .15, t = 3.16, p = .002 in Study 10; likelihood estimate: b = 5.44, SE = 1.78, t = 3.05, p = .002 in Study 9 and b = 5.21, SE = 2.12, t = 2.45, p = .014 in Study 10). Merely choosing the possibly correct ranges did not increase or decrease confidence relative to the Control condition (confidence rating: b = .09, SE = .13, t = .68, p = .496 in Study 9 and b = .04, SE = .15, t = .26, p = .793 in Study 10; likelihood estimate: b = 2.66, SE = 1.84, t = 1.45, p = .149 in Study 9 and b =

1.65, SE = 2.08, t = .79, p = .428 in Study 10). Most importantly, first choosing the possibly correct ranges and then constructing the belief distribution also increased confidence relative to the Control condition (confidence rating: b = .43, SE = .14, t = 3.14, p = .002 in Study 9 and b = .29, SE = .16, t = 1.85, p = .065 in Study 10; likelihood estimate: b = 5.26, SE = 1.82, t = 2.89, p = .004 in Study 9 and b = 4.73, SE = 2.24, t = 2.11, p = .035 in Study 10).[10] These results were consistent with our expectations. Overconfidence was increased by allocating probabilities to outcomes, but not by merely considering all possible outcomes.

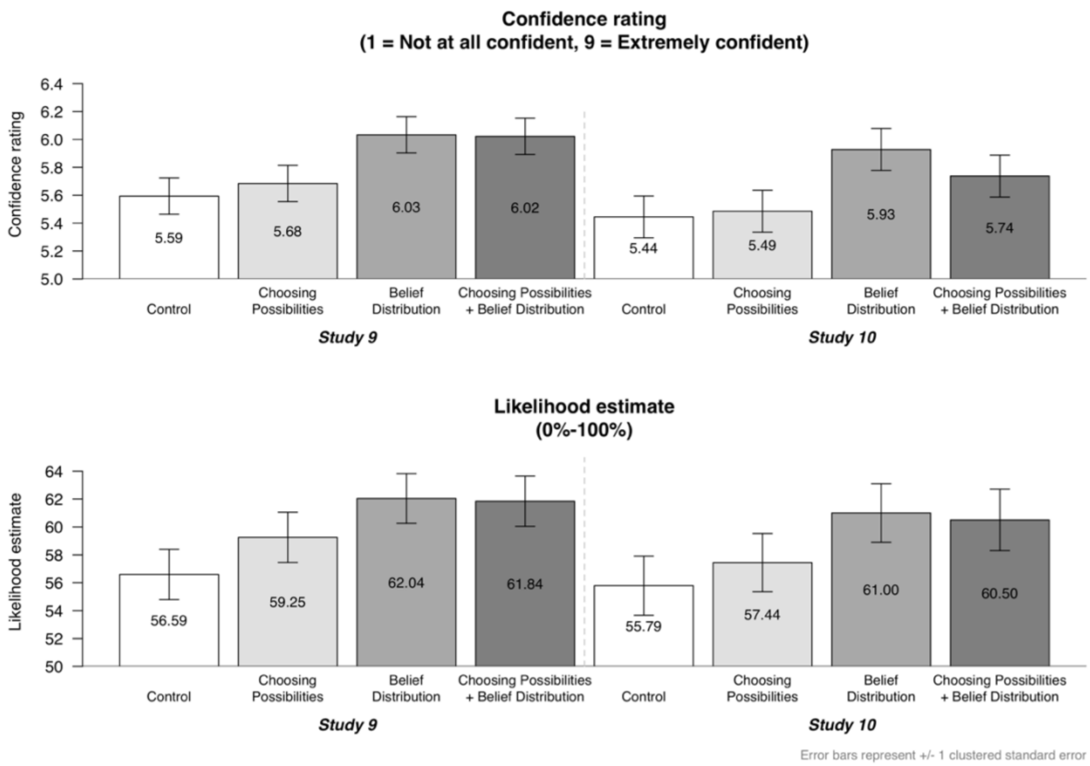

Figure 8. Studies 9 and 10 results

Note. Error bars represent +/-1 clustered standard error.

---

[10] It is worth noting that more than half of participants in the Choosing Possibilities + Belief Distribution condition (61.9% in Study 9, 65.0% in Study 10) allocated probabilities to outcomes that they did not choose as possible on the previous page. This is consistent with our claim in the beginning of the article that many participants feel compelled by the belief distribution task to allocate mass into more categories even if they do not consider these categories as possible, a fact that would artificially widen the 90% confidence intervals. We thank an anonymous reviewer for suggesting this analysis.

General Discussion

Prior research suggests that one can reduce overconfidence by eliciting a belief distribution over the entire range of outcomes (e.g., Haran et al., 2010; Moore, 2020). However, this conclusion was largely based on studies that measured overconfidence differently in different conditions and thus confounded the key manipulation with the measure of interest. In the present investigation, we provided cleaner tests of this important hypothesis. Specifically, we conducted 10 pre-registered studies in which we assessed confidence using the same face-valid measures across all conditions.

Contrary to what past research suggests, and to our own expectations, we found that providing a belief distribution usually *increases* people's confidence.[11] To see just how consistent this result is, we have plotted, in Figure 9, the standardized mean difference (i.e., Cohen's *d*) in confidence between the Belief Distribution condition and the Control condition for each item in Studies 2-10 (i.e., the studies that included a Control condition). The top panel shows the results for the confidence ratings and the bottom panel shows the results for the likelihood estimates. You can see that of the 46 comparisons presented in each figure, the overwhelming majority were positive, indicating that the Belief Distribution condition increased confidence relative to the Control condition. All of the few directionally negative effects came in a single study (Study 4), and none of those were statistically significant. We interpret these results as indicating that providing a belief distribution usually exerts a small but *positive* effect on overconfidence.

[11] It is worth reiterating that by increasing confidence, providing a belief distribution also increases *over*confidence in our studies. As we note in the preliminary analyses for Studies 1 and 2, participants were overconfident in every condition of every study. In addition, participants' calibration (as measured by the correlation between prediction accuracy and likelihood estimates) does not differ meaningfully across conditions (with full details reported in Supplement 4).

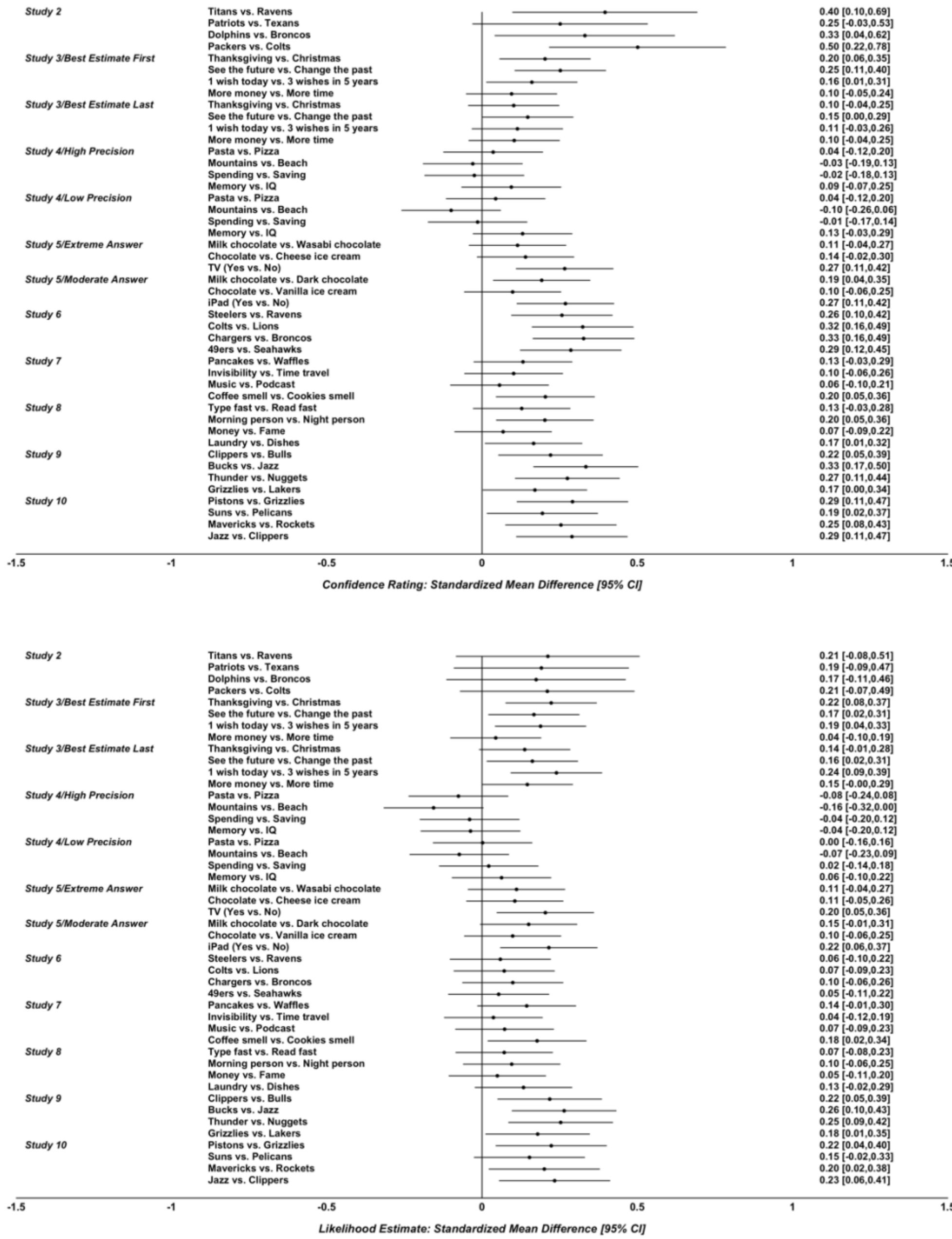

Figure 9. Forest Plots of the Standardized Mean Difference in Confidence Ratings (Top Panel) and Likelihood Estimates (Bottom Panel) Between the Belief Distribution Condition and the Control Condition (Belief Distribution Condition Minus Control Condition)

Note. A positive sign reflects that the Belief Distribution condition increased confidence compared to the Control condition; a negative sign reflects that the Belief Distribution condition reduced confidence compared to the Control condition.

Our 10 studies were diverse enough to suggest that this effect is fairly generalizable and robust. We observed it in two very different prediction domains, we observed it regardless of whether participants made their predictions before or after providing their belief distributions (see Studies 3, 9, and 10), and we observed it when correct answers were extreme and obvious or moderate and arguably less obvious (Study 5). In Study 4, we did not find any effect of providing a belief distribution on overconfidence, but we did observe that it did not seem to be moderated by whether the predictions were framed to be more or less precise.

Possible Explanations

In our later studies, specifically Studies 8-10, we discovered that belief distributions seem to increase overconfidence because the act of allocating probabilities to outcomes reinforces people's initial beliefs. We did not observe a similar effect when people merely considered which outcomes were possible. Why would allocating probabilities to outcomes increase people's confidence?

Further analyses of our data show that most participants allocated probabilities in a way that served to reinforce their predictions. For example, most participants allocated a greater than 50% probability to the three outcome intervals that were closest to their prediction, and many allocated as much as 75% or 90% to these three intervals (see Supplement 6). Moreover, as shown in Supplement 6, participants tended to express higher confidence when they allocated higher probabilities to the bins that were closest to their best estimate.

Just so you visualize this, we've constructed Figure 10, which displays the belief distributions provided by a random subset of 12 participants for an arbitrarily selected prediction question. As you can see, most participants provided distributions that assigned very high probabilities to outcomes that were close to their estimates, and low or no probabilities to outcomes that were far from their estimates. (You can also see that a few participants probably did not take this task very seriously.) Take Participant 33 in Figure 10 as an example. This participant gave an initial prediction of 25% and constructed a belief distribution with the majority of mass allocated to the range of 21-40%. Instead of realizing that other answers (e.g., 10% or 60%) might be possible, they may become more convinced in their initial predictions as a result of this process. Indeed, it makes sense that constructing a belief distribution that assigns such a large probability to forecasted or nearly-forecasted outcomes would serve to increase rather than decrease confidence in one's initial predictions.

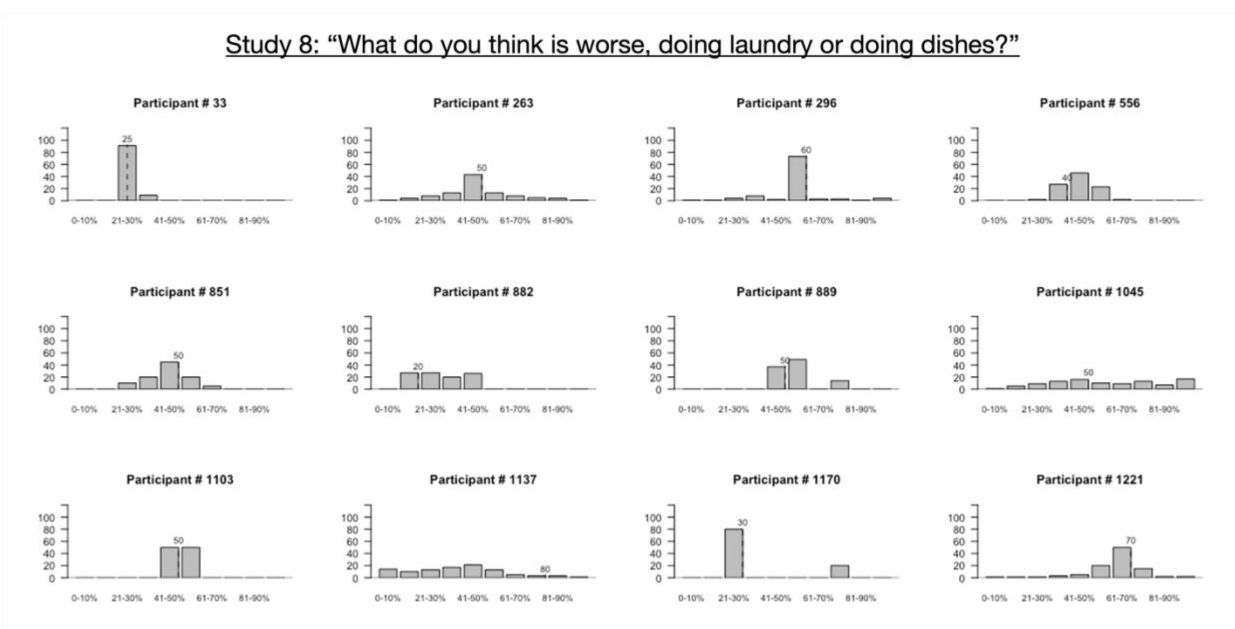

Figure 10. Examples of Belief Distributions Constructed by Individual Participants for One Item in Study 8.

Note. We arbitrarily selected one of the prediction items and plotted the belief distributions provided by 12 randomly selected participants in the dataset. Within each panel, the dashed line and number above it indicates the best estimate provided by the participant.

Note that this result – that people tend to allocate probabilities to outcomes in a way that reinforces their predictions – is not at odds with our earlier claim that many participants may also feel compelled by the belief distribution task to give 90% confidence intervals that are artificially wide. Both things can be simultaneously true: Participants can allocate most of the probabilities to outcomes that are very close to their forecasts, while at the same time allocating some probabilities to outcomes that they believe to be impossible (participant 296 in Figure 10 is a good example of this; also see Footnote 10.) Indeed, as described in Supplement 5, we do find that despite participants' tendency to allocate high probabilities to outcomes that are very close to their forecasts when constructing their belief distributions, the 90% confidence intervals imputed from those distributions are usually wider than the 90% confidence intervals that participants report when they are directly elicited.

Past demonstrations of the "dud-alternative effect" (Windschitl & Chambers, 2004; Windschitl & Wells, 1998) may suggest another reason why allocating probabilities might reinforce beliefs. Windschitl and Chambers (2004) found that when implausible alternatives (i.e., duds) were present, people judged a focal hypothesis to be more likely and were willing to bet more money on it.[12] Windschitl and colleagues proposed that when people make likelihood judgments without much deliberate effort, they compare evidence supporting the focal hypothesis against evidence supporting each of the alternatives. The existence of duds increases the number of favorable

[12] They did not, however, find this effect when likelihood judgments were elicited with numeric measures (e.g., when they asked participants to indicate the numeric likelihood of the focal outcome).

comparisons for the focal hypothesis and, therefore, increases its perceived likelihood. In a similar vein, it is possible that when participants are presented with the full range of possible outcomes, the presence of weak alternatives increases the perceived likelihood of the initial prediction. For example, when participants predict the percentage of respondents choosing chocolate ice cream over vanilla ice cream, the peripheral categories (e.g., 0-10%, 91-100%) would be perceived by most people as implausible. As the process of allocating probabilities provokes a search for evidence supporting each alternative, it might facilitate the realization that these categories are extremely weak alternatives and, as a result, strengthen people's initial beliefs.

Our findings are surprising given past literature on overconfidence and research touting the effectiveness of "consider the alternative" debiasing interventions. Past work shows that when prompted to take multiple viewpoints or consider a wider range of possible outcomes, people are less prone to stick to their initial judgments, and they become less overconfident (Hoch, 1985; Koriat et al., 1980; Lord et al., 1984; Walters et al., 2017). Because providing a belief distribution necessarily forces people to consider all possible outcomes, it is reasonable to expect it to decrease confidence. Contrary to what prior literature suggests, even asking people to consider all possibilities without assigning probabilities to each one (i.e., the Choosing Possibilities condition of Studies 8-10) did not consistently reduce confidence relative to a Control condition that merely provided a best estimate.

Indeed, our investigation suggests that merely considering all possible outcomes may not be enough to reduce overconfidence, simply because people may believe that many of these outcomes are unlikely. In two non-pre-registered, exploratory studies (Studies S4 and S5, reported in Supplement 11), we asked participants to make a prediction, indicate how confident they were in that prediction, provide a belief distribution, and then rate their confidence again. We then asked the minority of participants who changed their confidence ratings after providing a belief distribution to tell us why. Most participants who increased confidence after providing a belief distribution indicated that the belief distribution task made them realize that their prediction was more likely and/or that alternative outcomes were not very plausible. On the other hand, many participants who reduced confidence said that considering the entire distribution of outcomes made them realize there were other possible outcomes that they did not consider before. Collectively, these open-ended responses suggest that those who deemed the non-forecasted alternatives as more plausible were more likely to reduce their confidence after completing the belief distribution task. This suggests that for an intervention to effectively reduce overconfidence, it may need to actively convince people that the non-forecasted outcomes are more likely than participants previously believed. Mere consideration of these outcomes is not sufficient.

Future Directions

Our investigation leaves some open questions. First, one plausible explanation for why allocating probabilities increases overconfidence implicates the observation that participants tend to construct belief distributions by assigning large probabilities to forecasted or nearly forecasted outcomes (as Figure 10 and Supplement 6 illustrate). According to this account, participants see themselves allocating most probabilities to outcomes closely around their best estimate and thus are more likely to believe that they were right all along. Although we replicated our results across sports predictions and preference predictions, it is possible that these happen to be the prediction domains where people tend to construct highly concentrated belief distributions, rendering our results more likely to occur. There might be occasions in which the opposite occurs. That is, in prediction domains where people are more likely to construct dispersed belief distributions (i.e., assigning probabilities more evenly across all categories), perhaps constructing belief distributions will serve to reduce people's confidence. Future research could investigate this possibility and further explore the mechanism through which constructing belief distributions increases overconfidence.

Second, in our research we recruited online (nonexpert) participants in all our studies and thus our research cannot speak to how belief distribution elicitations would influence experts' overconfidence.[13] Although we believe that online participants' results serve as a useful benchmark for those of experts', the two samples likely differ in many aspects, such as their baseline confidence and the range of outcomes they consider by default. Extending the current research to expert samples would have important implications, especially since the belief distribution elicitations have been used in geopolitical forecasting tournaments (Moore et al., 2017) and have been proposed as a useful tool for business forecasting (Haran & Moore, 2014). We look forward to future research that attempts to shed light on whether our findings generalize to expert judgments.

Third, exploratory analyses of our data show that the effect size comparing the Belief Distribution condition and the Control condition declines over the course of several predictions. That is, providing a belief distribution led to the largest increase in confidence for the first prediction, and the magnitude of confidence increase was reduced for the second, third, and fourth predictions (see Supplement 7). We do not know why this happens. Perhaps it is merely a methodological artifact, reflecting less participant engagement for items that came later in the survey. Or perhaps it is psychologically meaningful, indicating that, for example, belief

[13] In our sports prediction studies (Studies 1, 2, 6, 9, and 10), we collected participants' domain knowledge for exploratory purposes. We tested the interaction between the knowledge score and the belief distribution condition. Across five studies, this interaction was significant only in Study 9 (b = -.25, clustered SE = .06, p < .001 for confidence ratings, b = -3.35, clustered SE = .95, p < .001 for likelihood estimates), where the belief distribution manipulation increased confidence more among participants with lower knowledge scores.

distributions exert reduced effects on confidence once participants become more familiar with them. Future research could try to find out.

Context

Researchers in psychology, economics, and related fields have expressed a growing interest in assessing and analyzing people's entire belief distributions, as doing so potentially offers a more detailed understanding of what participants believe. Meanwhile, identifying interventions to reduce overconfidence has been a decades-long enterprise. The current research was inspired by both streams of research and specifically by an important claim in past research that constructing belief distributions can reduce overconfidence. While past research was confounded, we hoped that providing a clean test of this claim would not only elucidate the interpretation of past findings but also help us better understand how people construct belief distributions and shed light on the origins of overconfidence. The current work shows that constructing belief distributions – a method proposed to reduce overconfidence – can inadvertently exacerbate overconfidence. We look forward to future research that investigates why this occurs, and that potentially unearths conditions under which constructing belief distributions truly reduces overconfidence. This research was conducted as part of the first author's dissertation, which examines how belief distributions are constructed and how they influence people's beliefs. It also extends the second author's research on understanding the biases that plague people's predictions (e.g., Kelly & Simmons, 2016; Simmons & Massey, 2012; Simmons & Nelson, 2006).

References

Alpert, M., & Raiffa, H. (1982). A progress report on the training of probability assessors.

André, Q., Reinholtz, N., & De Langhe, B. (2022). Can consumers learn price dispersion? evidence for dispersion spillover across categories. *Journal of Consumer Research*, *48*(5), 756-774.

Arkes, H. R., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes*, *39*(1), 133-144.

Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. *Organizational Behavior and Human Decision Processes*, *49*(2), 188-207.

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychological Science, 31*(10), 1302-1314.

Goldstein, D. G., & Rothschild, D. (2014). Lay understanding of probability distributions. *Judgment and Decision Making*, *9*(1).

Haran, U., Moore, D. A., & Morewedge, C. K. (2010). A simple remedy for overprecision in judgment. *Judgment and Decision Making*, *5*(7), 467.

Haran, U., & Moore, D. A. (2014). A better way to forecast. *California Management Review*, *57*(1), 5-15.

Higginbotham, A. (2019). *Midnight in Chernobyl: The untold story of the world's greatest nuclear disaster*. Random House.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 719.

Hofman, J. M., Goldstein, D. G., & Hullman, J. (2020, April). How visualizing inferential uncertainty can mislead readers about treatment effects in scientific results. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1038.

Kelly, T. F., & Simmons, J. P. (2016). When does making detailed predictions make predictions worse?. *Journal of Experimental Psychology: General*, *145*(10), 1298.

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, *32*, 385-418.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes*, *79*(3), 216-247.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 107.

Langnickel, F., & Zeisberger, S. (2016). Do we measure overconfidence? A closer look at the interval production task. *Journal of Economic Behavior & Organization*, *128,* 121-133.

Lewis, M. (2011). *The big short: Inside the doomsday machine*. Penguin UK.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. *Decision Making and Change in Human Affairs*, 275-324.

Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance*, *26*(2), 149-171.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: a corrective strategy for social judgment. *Journal of Personality and Social Psychology*, *47*(6), 1231.

McLean, B., & Elkind, P. (2013). *The smartest guys in the room: The amazing rise and scandalous fall of Enron*. Penguin.

Moore, D. A. (2020). *Perfectly confident: How to calibrate your decisions wisely*. HarperCollins.

Moore, D. A. (2022). Overprecision is a property of thinking systems. *Psychological Review*.

Moore, D. A., Carter, A. B., & Yang, H. H. (2015). Wide of the mark: Evidence on the underlying causes of overprecision in judgment. *Organizational Behavior and Human Decision Processes*, *131*, 110-120.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502.

Moore, D. A., Swift, S. A., Minster, A., Mellers, B., Ungar, L., Tetlock, P., ... & Tenney, E. R. (2017). Confidence calibration in a multiyear geopolitical forecasting competition. *Management Science*, *63*(11), 3552-3565.

Prims, J. P., & Moore, D. A. (2017). Overconfidence over the lifespan. *Judgment and Decision Making*, *12*(1), 29.

Reinholtz, N., Fernbach, P. M., & De Langhe, B. (2021). Do People Understand the Benefit of Diversification?. *Management Science*, *67*(12), 7322-7343.

Ren, Y., & Croson, R. (2013). Overconfidence in newsvendor orders: An experimental study. *Management Science*, *59*(11), 2502-2517.

Simmons, J. P., & Massey, C. (2012). Is optimism real?. *Journal of Experimental Psychology: General*, *141*(4), 630.

Simmons, J. P., & Nelson, L. D. (2006). Intuitive confidence: choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General*, *135*(3), 409.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299.

Soll, J. B., Palley, A., & Klayman, J. (2019). Overconfidence in probability distributions: People know they don't know but they don't know what to do about it.

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., & Burgman, M. (2010). Reducing overconfidence in the interval judgments of experts. *Risk Analysis: An International Journal*, *30*(3), 512-523.

Teigen, K. H., & Jørgensen, M. (2005). When 90% confidence intervals are 50% certain: On the credibility of credible intervals. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *19*(4), 455-475.

Walters, D. J., Fernbach, P. M., Fox, C. R., & Sloman, S. A. (2017). Known unknowns: A critical determinant of confidence and calibration. *Management Science*, *63*(12), 4298-4307.

Williams, J. J., & Mandel, D. R. (2007). Do evaluation frames improve the quality of conditional probability judgment?. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 29, No. 29).

Windschitl, P. D., & Chambers, J. R. (2004). The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 198.

Windschitl, P. D., & Wells, G. L. (1998). The alternative-outcomes effect. *Journal of Personality and Social Psychology*, *75*(6), 1411